

LncRNA-mRNA co-expression analysis discovered the diagnostic and prognostic biomarkers and potential therapeutic agents for myocardial infarction

Xiaocong Zhang^{1,2,*}, Ziqi Chen^{2,*}, Jiabin Zang², Chun Yao², Jian Shi², Ruqiong Nie¹, Guifu Wu^{2,3,4}

¹Department of Cardiology, Sun Yat-Sen Memorial Hospital of Sun Yat-Sen University, Guangzhou, China

²Department of Cardiology, The Eighth Affiliated Hospital of Sun Yat-Sen University, Shenzhen, Guangdong, China

³Guangdong Innovative Engineering and Technology Research Center for Assisted Circulation, Shenzhen, China

⁴NHC Key Laboratory of Assisted Circulation, Sun Yat-Sen University, Guangzhou, Guangdong, China

*Equal contribution

Correspondence to: Guifu Wu, Ruqiong Nie; email: wuguifu@mail.sysu.edu.cn, nierq@mail.sysu.edu.cn

Keywords: myocardial infarction, lncRNAs, diagnostic models, support vector machine, weighted gene co-expression network analysis

Received: November 16, 2020

Accepted: February 9, 2021

Published: March 5, 2021

Copyright: © 2021 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Currently, the role of lncRNA in myocardial infarction (MI) is poorly understood. 17 co-expression modules were determined, specifically, the greenyellow, saddlebrown, grey60, royalblue, lightgreen, white, and pink modules were specifically expressed in the acute phase of MI, and brown, darkred, and royalblue, while greenyellow modules were specifically expressed in MI compared with CAD. 12 time-dependent of lncRNA/mRNA clusters with consistent expression trends were also identified. MI-associated modules were mainly enriched to immune, cell cycle, and metabolic pathways. We further obtained a network of 1816 lncRNA-mRNAs with higher expression correlations among these lncRNAs by analyzing the topological properties of the network. Herein, lncRNA RP11-847H18.2 and KLHL28, SPRTN, and EPM2AIP1 were determined as gene markers specifically expressed in MI, and they demonstrated a high predictive performance for MI diagnosis and prognosis. Three drugs, namely, Calcium citrate, Calcium Phosphate, and Calcium phosphate dihydrate, were identified as potential precursors of MI. Finally, gene and lncRNA diagnostic models were developed based on these genes and lncRNAs, with their AUCs averaged above 0.89 in both training and validation datasets. The findings of this study improve the diagnosis and prognosis of MI and personalized treatment of MI.

INTRODUCTION

Myocardial infarction (MI) is an acute cardiovascular disease with high mortality and disability [1]. On the basis of atherosclerotic stenosis of coronary arteries, MI is an acute myocardial necrosis caused by plaque rupture and sudden obstruction of the coronary artery lumen by some stimuli, resulting in continuous and severe hypoxia-ischemia of myocardial tissues (the innervation

of the infarcted vessel). After the occurrence of MI, apoptotic cascade will be activated and cardiomyocytes will become necrotic, due to persistent hypoxia and ATP deficiency in cardiomyocytes. Necrotic cardiomyocytes activate the immune system, subsequently leading to the excessive production of inflammatory response [2]. Therefore, it is important to explore the molecular mechanisms of MI pathogenesis and identify diagnostic and prognostic biomarkers and therapeutic targets.

Long noncoding RNAs (lncRNAs) are a class of RNA molecules with a transcript length of ≥ 200 nt. lncRNAs do not encode proteins, but regulate gene expression at epigenetic, transcriptional and post-transcriptional levels through gene imprinting, chromatin reconstruction, regulation of cell cycle, variable shear, and mRNA regulation, participating in various functional processes such as cell metabolism, growth, differentiation, apoptosis and death [3, 4]. lncRNAs play important regulatory roles in the pathological processes of AMI, including in cardiomyocyte apoptosis, inflammatory response, angiogenesis, fibrosis repair and cardiac remodeling. Studies proved that some differentially expressed lncRNAs, for instance, aHIF [5], ANRIL [5], KCNQ10T1 [6], MIAT [7], ZFAS1 [8] and CDR1AS [9], have critical functions in acute myocardial infarction (AMI) and demonstrated potentials of serving as myocardial infarction-specific biomarkers or key lncRNAs in the development regulation of AMI. Moreover, Liu Cuiyun et al [10] found that lncRNA CAIF inhibits myocardial autophagy and myocardial infarction by blocking the expression of p53-regulated myocardin. These previous findings indicated that lncRNAs have key functions in AMI, but the mechanism of lncRNA regulation in myocardial infarction is far from clear. Therefore, to investigate the specific role of lncRNAs in myocardial infarction and the related regulatory mechanisms are of high significance.

In this study, 12 time-dependent lncRNA/mRNA clusters with consistent expression trends and 1816 lncRNA-mRNA networks were obtained from a co-expression network of mRNA and lncRNA established. We also determined specific expressions of lncRNARp11-847H18.2 and three genes (KLHL28, SPRTN, EPM2AIP1), which showed strong potential in MI diagnosis. Finally, a lncRNA/mRNAs-based diagnosis model was constructed with the corresponding lncRNA and mRNAs.

RESULTS

Identification of modules for different stages of specific expression

The flow chart of our execution is shown in Figure 1. The "WGCNA" package in R was used to group Gene/lncRNA with similar expression patterns into modules by average-linkage hierarchical clustering. In this study, with $\beta = 16$ (scale-free $R^2 = 0.87$) power serving as the soft threshold to ensure a scale-free network (Figure 2A, 2B). A total of 17 modules were identified (Figure 2C). The correlation between each module and feature was calculated respectively. We found that the acute phase (MI_S1) at four time points

was highly correlated with multiple modules expressed in MI_S1 ($p < 1e-4$), which were Greenyellow, SaddleBrown, Grey60, RoyalBlue, Lightgreen, White, and Pink (Figure 2D). However, the correlation between the other three time points and each module was weak, indicating that the greatest transcriptome changes in MI occurred at the acute phase. Compared with the four time points, CAD also showed a high correlation with several modules ($p < 1e-4$), such as Brown, Darkred, Royalblue, Greenyellow, etc., In addition, we obtained data sets GSE57338 for Failing Hearts and controls from the GEO database, and calculated the relationship between eigenvectors of each module and healthy controls using the same method (Figure 2C). The results showed that the healthy control samples were highly correlated with multiple modules, similar to MI_S1, but opposite to CAD, suggesting that multiple transcriptome changes take place in the process of MI compared with the normal control group.

Functional dimensions of MI related modules and expressed timing analysis

KEGG functional enrichment analysis was performed for a better understanding of the functional implications of the nine MI-related modules. When $FDR < 0.01$, genes were enriched to a total of 56 modules (Figure 3A), but the pathways these modules enriched to showed less intersection, suggesting that genes in different co-expressed modules may be involved in different biological processes in time or space. We also observed that the genes in these modules were not only related to multiple viral infection pathways such as Human T-cell virus 1 infection, Pathogenic *Escherichia coli* infection, Yersinia infection, and Human Syractomiae infection, but also to Th1 and Th2 cell differentiation, Th17 cell differentiation, T cell receptor signaling pathway, IL-17 signaling pathway and some other immune pathways. In addition, these genes were enriched into cell cycle, basal transcription factors, RNA transport, cellular senescence, apoptosis and other cell proliferation and apoptosis pathways. These results suggested that MI is a complex and systematic process involving cell proliferation, apoptosis, and immunity. Considering that MI_S2\S3\S4 did not show specific relation to co-expression module, we analyzed the patterns of gene and lncRNA expressions at different time points based on time series, and determined 12 expression patterns (Figure 3B). For example, Cluster 1 showed increased expression at MI_S2 and remained at a high level since then, while Cluster 11 demonstrated an opposite pattern to that of Cluster 1. Moreover, the expression of Cluster 3 increased continuously with time, while that of Cluster10 was the opposite. The distribution statistics of genes and lncRNAs in each Cluster (Supplementary Figure 1A) showed that the proportion of lncRNAs in

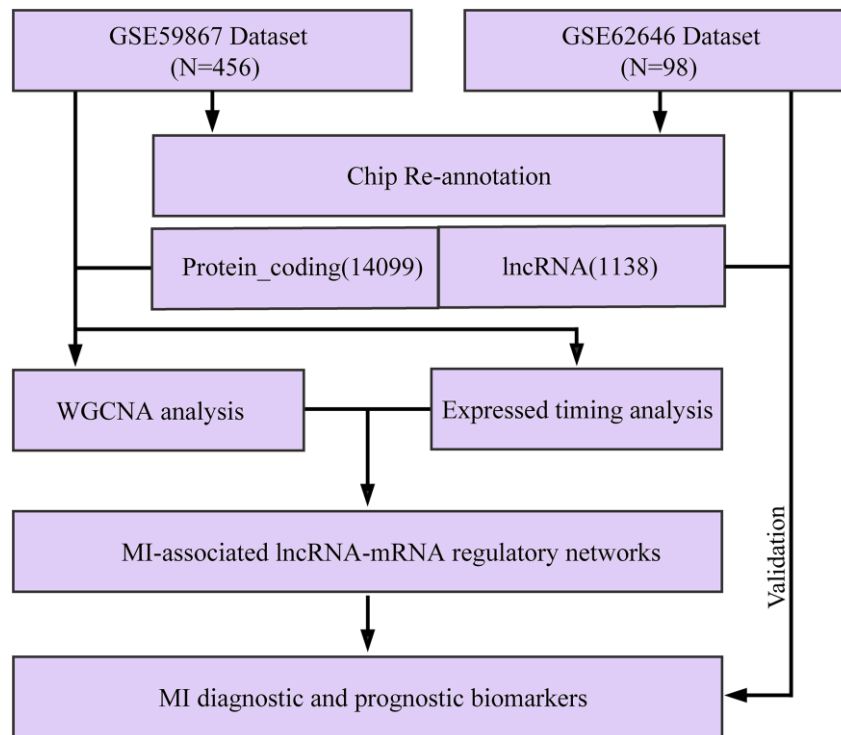


Figure 1. Workflow diagram.

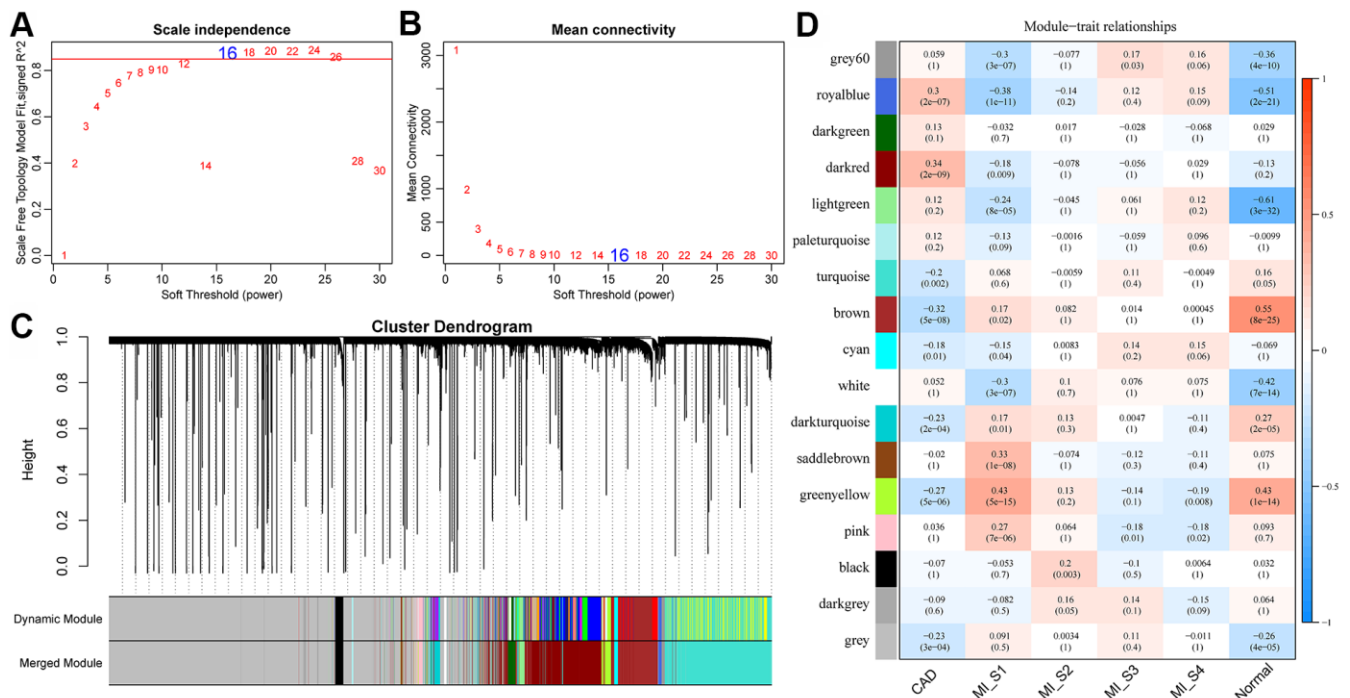


Figure 2. Construction of weighted gene co-expression network and identification of disease-related modules. (A, B) Determination of soft-thresholding power in the weighted gene co-expression network analysis (WGCNA). (A) Analysis of the scale-free fit index for various soft-thresholding powers (β). (B) Analysis of the mean connectivity for various soft-thresholding powers. (C) Dendrogram of all differentially expressed genes/lncRNAs clustered based on a dissimilarity measure (1-TOM). (D) Correlation distribution of feature vectors of each module with four time points and CAD.

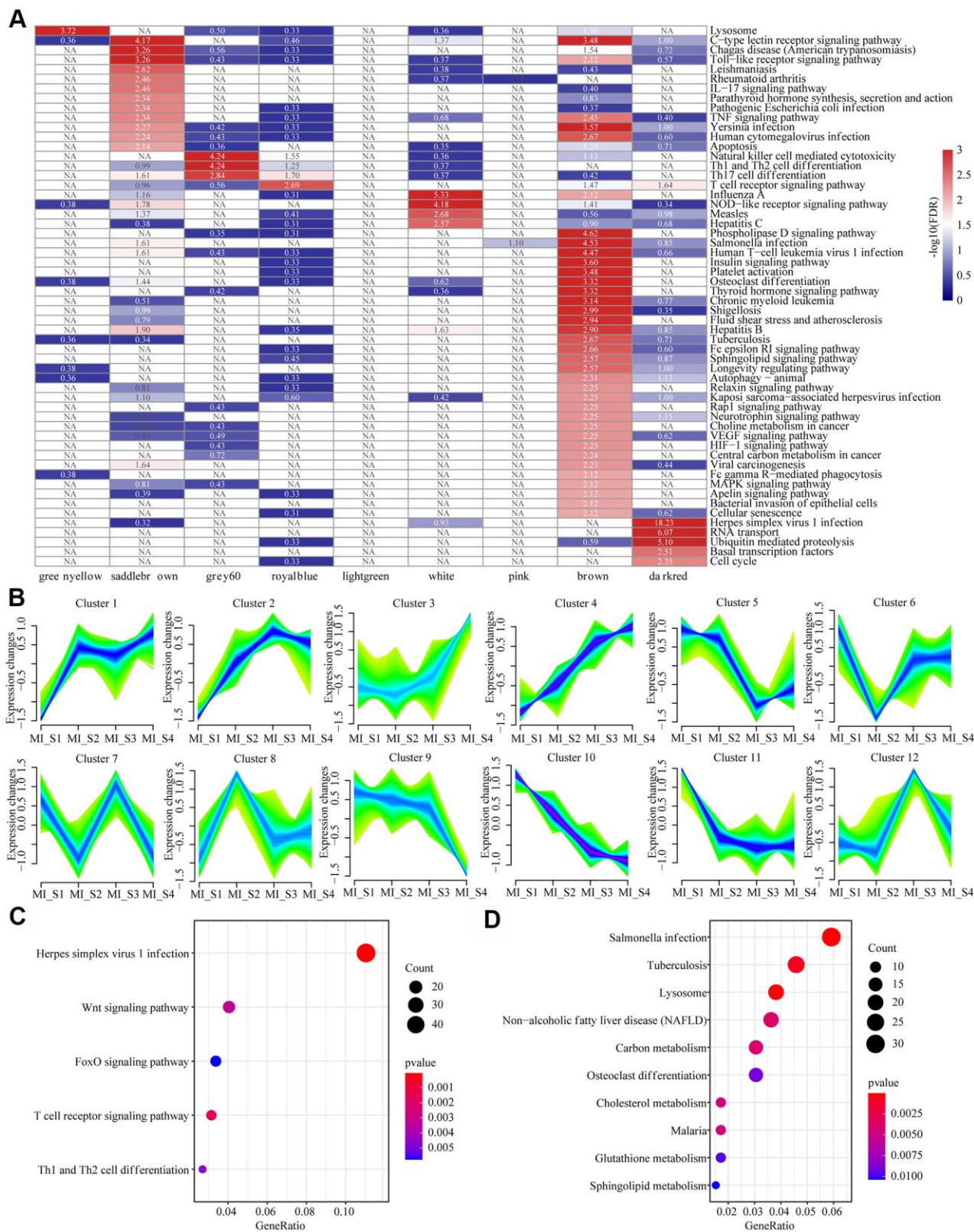


Figure 3. Functional enrichment analysis of MI related modules and timing analysis of MI gene expression. (A) Heatmap of 9 modules enriched to KEGG Pathway, color indicates \log_{10} (FDR), NA indicates unenriched. **(B)** Time series analysis of 12 expression patterns of gene expression at different time points. **(C)** Sustained expression of elevated gene sets enriched to the KEGG pathway at different time points. **(D)** The KEGG pathway was enriched for a set of genes with decreased expression at different time points. The different colors indicate enrichment significance and the dot size indicates the number of enriched genes.

all the expression patterns were low, which was consistent with the background distribution characteristics dominated by PCG. These expression patterns indicated that the transcriptome was changing continuously at different time points in MI, and these changes may lead to different disease progression. For KEGG pathway enrichment analysis, we selected the gene set showing continuously increased expression (Cluster 4) and the one with continuously decreased expression (Cluster 10). We observed that Cluster 4 was mainly enriched to the Wnt Signaling pathway, FoxO signaling pathway, T cell signaling pathway, and other signaling pathways (Figure 3C), and Cluster 10 was mainly enriched in Sphingolipid metabolism, glutathione metabolism, cholesterol metabolism, carbon metabolism and other metabolic pathways (Figure 3D). This suggested that the immune-related pathways of MI patients are gradually enhanced, by contrast, the carbon metabolism-related pathways are weakened over time.

Identification of MI-associated lncRNA-mRNA regulatory networks

We examined the distribution of genes and lncRNAs in MI-associated modules and two continuously changing expression patterns, and observed a significantly low proportion of lncRNAs in seven of these modules ($FDR < 0.05$), suggesting that lncRNAs may be indirectly involved in MI development through multiple regulatory pathways. Based on this, a new computational method was developed to identify the lncRNA-mRNA interaction network in MI. This was achieved by integrating paired expression profiles of genes/lncRNAs from disease-related co-expression modules into gene expression datasets according to the regulatory interactions among mRNAs, lncRNAs and miRNAs. Here we determined a network of 5320 lncRNA-mRNA interactions incorporating a total of 510 mRNAs and 154 lncRNAs (Table 1). The mRNAs and lncRNAs in the network showed significantly different degrees of distributions, with the average degree of lncRNAs greater than that of mRNA, suggesting that lncRNAs are more likely to be the hub nodes in the network (Figure 4A). The distribution of mRNA-lncRNA correlations was further analyzed, and the results demonstrated that there was a significant correlation between the distribution of between lncRNAs and mRNAs (Figure 4B). This suggested that the association between lncRNAs and mRNAs may have stronger interactions in MI than other gene correlations. Therefore, the intersection of different mRNA-lncRNA correlations were acted as the threshold, we further identified lncRNAs/mRNAs with correlations higher than 0.3 in the interaction network as MI-associated lncRNA-mRNAs, resulting in a network of 1816 lncRNA-mRNA interactions incorporating 417 genes and 112 lncRNAs

(Figure 4C). Further analysis detected several nodes with the highest and lowest degrees of distribution consistent with the biological network characteristics (Figure 4D). In addition, the network showed a median distribution (Figure 4E), the same near-centrality distribution (Figure 4F), and eigenvector centrality distribution (Figure 4G). All these results indicated that MI-associated lncRNA-mRNA is a canonical biological regulatory network centered on lncRNAs.

Functional enrichment analysis of MI-associated lncRNAs-mRNAs network

To observe the function of the MI-associated lncRNA-mRNA network, the genes in the network were subjected to functional enrichment analysis of GO and KEGG pathway. The analytical data demonstrated that these genes were mainly enriched to autophagy, insulin resistance and some other pathways (Figure 5A). Previous studies found that autophagy in blood cardiomyocytes can provide the energy necessary for cell survival by removing disordered organelles or senescent proteins [11]. For example, NR4A2 knock-down exacerbates cardiomyocyte apoptosis, and upregulation of NR4A2 is regarded as an adaptive response to ischemia-induced cardiomyocyte apoptosis [12]. Insulin resistance is considered as a key risk factor for adverse metabolic and cardiovascular diseases [13]. In the heart, insulin receptor substrates (IRS) are key regulators of the insulin signaling pathway. Under the activation of insulin receptor (INSR), IRS promotes glucose transporter protein 4 (SLC2A4, also known as GLUT4) translocation and initiates cell survival pathways in cardiomyocytes [14], in addition, reduced insulin sensitivity in ischemic myocardium contributes to defective IRS1 function [15, 16]. These findings demonstrated that the lncRNA-mRNA network may be involved in the development of MI through regulating autophagy and insulin resistance. Moreover, these genes were also found to be enriched in a variety of metabolic and decomposition processes (Figure 5B). The metabolic process of degrading damaged proteins or organelles is the main mechanism of autophagy. The enrichment results of cell components showed that these genes were enriched into a variety of complexes related to transcriptional regulation and histone modification (Figure 5C). In addition, molecular functions were mainly enriched in multiple RNA binding, nucleosome binding and modification-dependent protein binding (Figure 5D), suggesting that these genes are also involved in chromatin modification and regulation. Epigenetic mechanism plays a key role in the regulation of mammalian gene expression. Differential histone modification represents a typical epigenetic mechanism, meaning that epigenetic modification also participates in the occurrence and development of MI.

Table 1. Genes and lncRNAs in MI-related modules or expression patterns.

Tag	LNC	PCG	p.value	FDR
brown	31	1211	7.40E-16	6.66E-15
darkred	57	1974	2.69E-22	2.69E-21
greenyellow	5	248	2.31E-05	0.000162
grey60	7	104	0.268219	0.536437
lightgreen	4	145	0.003388	0.016941
pink	4	154	0.00197	0.011819
royalblue	0	76	NA	NA
saddlebrown	4	41	0.563794	0.563794
white	1	56	0.011879	0.047516
Cluster4	86	1264	0.04585	0.13755
Cluster10	46	1170	1.02E-08	8.14E-08

MI diagnostic and prognostic biomarkers were identified by lncRNA-mRNA network mining

In order to examine the diagnostic performance of the lncRNAs and mRNAs in the network, linear discriminant analysis was applied to classify and predict each lncRNA, mRNA, and lncRNA-mRNA pair in the network, respectively, and their prediction accuracy distribution was calculated accordingly (Figure 6A). We observed that lncRNA, mRNA and lncRNA-mRNA presented similar classification accuracy, with an average accuracy of more than 0.6, which verified the diagnostic performance of the nodes in the lncRNA-mRNA network. In addition, we also investigated whether early changes in gene expression can also predict disease prognosis and distinguish patients with HF after MI from those without HF. Here, the classification/prediction accuracy distribution of each lncRNA and mRNA as well as lncRNA-mRNA pair in the network were evaluated in the same way (Figure 6B), and their average prediction accuracy was found to be greater than 0.55. Noticeably, the prediction accuracy of lncRNA-mRNA was significantly higher than that of lncRNA and mRNA alone. We also observed that the classification accuracy distribution of lncRNA-mRNA presented four peaks, indicating the existence of a variety of different lncRNA-mRNA combinations. Therefore, lncRNA-mRNA combinations in the peak with the highest accuracy were selected as the final candidate markers. Based on this, lncRNA-mRNA pairs with a diagnostic and prognostic accuracy greater than 0.7 were selected as candidate markers. In addition, we also found that nodes with higher network degree, intermediate number centrality, near centrality and feature vector centrality in biological networks were more likely to be the core nodes in the network regulation process. Thus, the nodes with the top 10% of network degree, medium centrality, near centrality, and eigenvector centrality were determined as the hub nodes

of the network. The intersection of them and candidate markers contained a total of one lncRNA (RP11-847H18.2) and three genes (KLHL28, SPRTN, and EPM2AIP1) (Figure 6C). These four molecules combined showed a high classification performance with high network importance, and can therefore be used as diagnostic and prognostic markers of MI. In addition, we also analyzed the relationship between the three genes and drugs, obtained the protein interaction network composed of drug target genes and our three genes through STRING database, constructed the drug-target gene-disease-specific gene interaction network, and determined the shortest distribution path of drugs to these three disease-specific genes in the network (Figure 6D). We observed that the average shortest path of most drugs was 10, while that of three drugs was only 4, suggesting that these three drugs may be effective for MI treatment (Figure 6E).

Construction of the MI diagnostic model and testing of the model

GSE59867 was considered as a Train dataset, to balance the proportion of control and disease groups, we randomly selected 46 samples from 73 patients in MI_S1 from the control group to form a cohort with 46 samples as a training set (TrainSet), and another 46 samples were selected from MI_S2, MI_S3, and MI_S4 to form a cohort with the control group, respectively. The three datasets used as internal validation sets were TestSet1, TestSet2, and TestSet3. Similarly, GSE62646 served as an external validation set, and equal proportions of MI samples corresponding to the three follow-up time points in the GSE62646 dataset and 14 control samples were gathered together to form three validation datasets, namely, ValidationSet1, ValidationSet2, and ValidationSet3. In the training set, one lncRNA and three genes served as features, and their expression profiles were obtained to construct a

support vector machine (SVM) classification model, which was tested by a tenfold cross-validation method. The AUC of the training set was 0.97 with 90% classification accuracy, 0.89 with 83% classification accuracy in TestSet1, 0.85 with 77% classification accuracy in TestSet2, and 0.85 with 74% classification accuracy in TestSet3 (Figure 7A, 7B). We further

applied the model to the external validation set GSE62646, and obtained an AUC of 0.91 and classification accuracy of 79% in ValidationSet1, an AUC of 0.94 and classification accuracy of 79% in ValidationSet2, and an AUC of 0.86 and classification accuracy of 75% in ValidationSet3 (Figure 7C, 7D). These results indicated that the diagnostic and

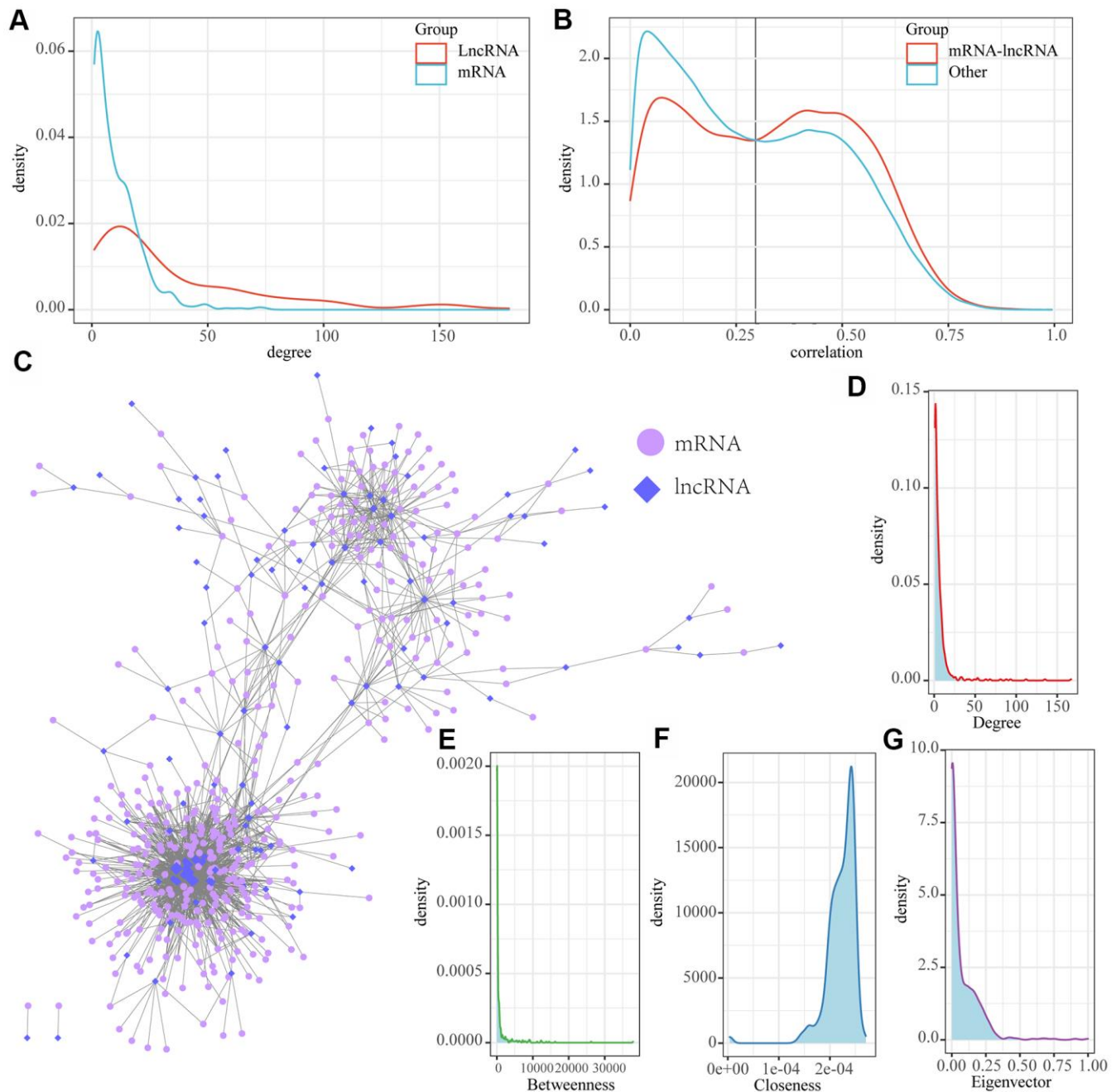


Figure 4. MI-associated lncRNA-mRNA regulatory networks. (A) Distribution degree of lncRNA and mRNA nodes in lncRNA-mRNA interaction networks; (B) Distribution of lncRNA and mRNA correlations in lncRNA-mRNA regulatory networks and lncRNA-mRNA correlations in non-lncRNA-mRNA networks; (C) MI-associated lncRNA-mRNA regulatory networks. (D) Distribution degree of the network. (E) Median centrality distribution of the network. (F) The near-central distribution of the network. (G) The eigenvector centrality distribution of the network.

predictive model constructed in this study can effectively distinguish MI patients from CAD controls, and that the one lncRNA and three genes can be used as reliable biomarkers for MI diagnosis.

DISCUSSION

The aim of this study was to identify potential diagnostic and prognostic biomarkers from the GEO database for myocardial infarction (MI). In this study, a candidate lncRNA and three mRNAs as potential biomarkers of MI after performing an exhaustive literature review. Our linear discriminant analysis showed that the lncRNA-mRNA pair exhibited a better performance in predicting prognosis than using individual an lncRNA or mRNA. Our target prediction and pathway analysis found that the lncRNA-mRNA network typically regulated autophagy and insulin resistance pathways. The results of drug-target gene interaction data proved that three drugs (Calcium phosphate, Calcium citrate and Calcium phosphate

dihydrate) had strong potentials in protecting patients against MI. Furthermore, we selected the largest molecular weight drug (DB11093: Calcium citrate) and evaluated the relationship with the three core genes in the network (KLHL28, SPRTN, EPM2AIP1) by molecular docking (Supplementary Figure 1B), the docking results with SPRTN, KLHL28 and EPM2AIP1L showed that the docking scores of the ligand with three receptor proteins were -5.4 kcal/mol, -5.8 kcal/mol and -5.0 kcal/mol, respectively. The docking scores of this ligand with the receptor all reached 5.0 kcal/mol at a small molecular weight loudness, indicating that the molecule has some binding affinity to the three proteins. In addition, DB11093 binds tightly to the zinc ion in the SPRTN protein in the form of a three-foot chelate. The compound can also be found to interact with VAL381, ALA427 and PHE324 in the protein by hydrogen bonding when DB11093 binds to KLHL28 protein. And when DB11093 binds to EPM2AIP1L protein, it can also form hydrogen bonding interactions with ARG343, ASN334 and

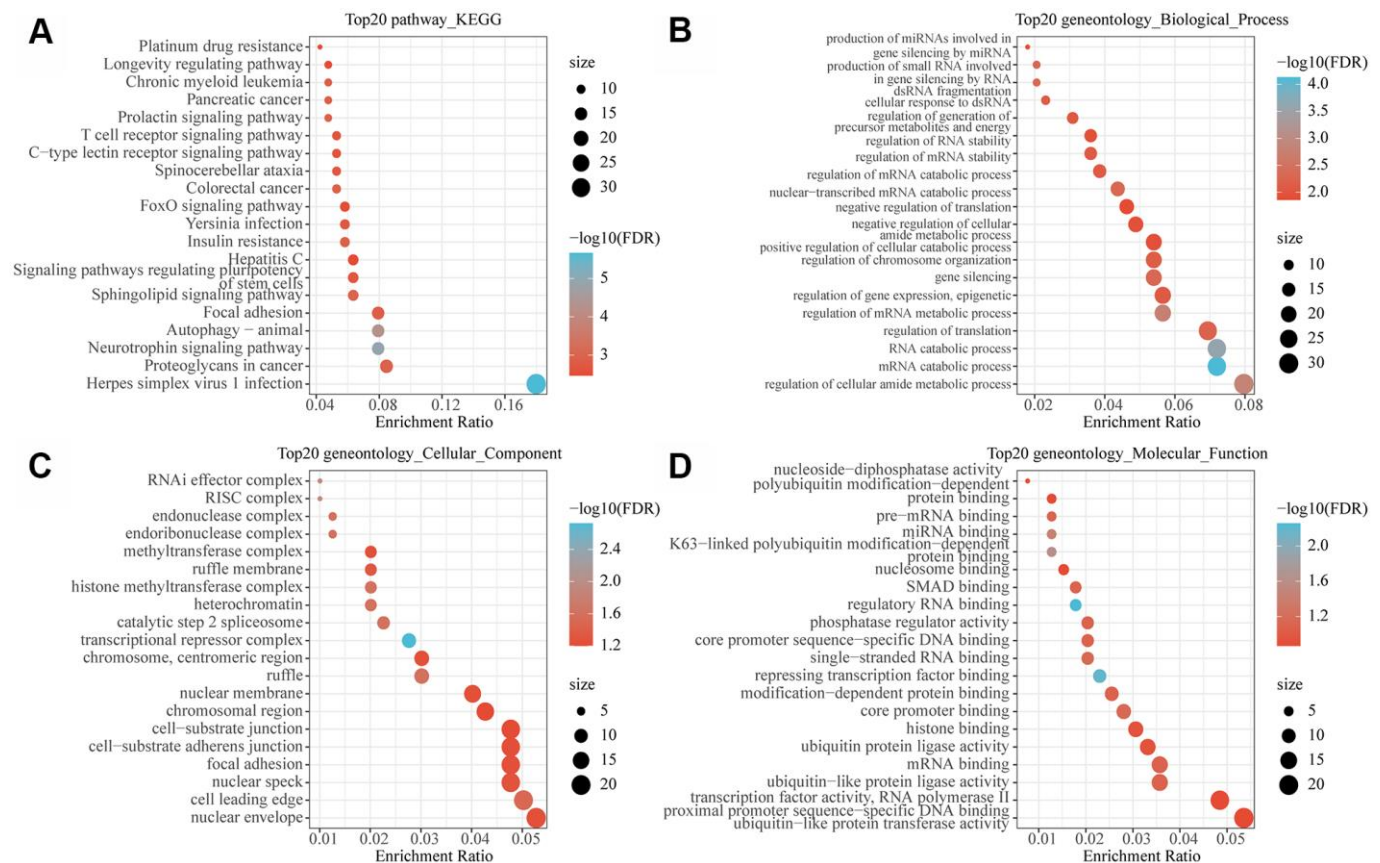


Figure 5. Functional enrichment analysis of MI-related lncRNA-mRNA networks. (A) The 20 most significant KEGG pathways enriched by MI-related lncRNA-mRNA network. (B) The 20 most significant GO biological processes enriched by MI related lncRNA-mRNA network. (C) The 20 most significant GO biological processes enriched by the MI-related lncRNA-mRNA network. (D) The 20 most significant GO molecular functions enriched by the MI-related lncRNA-mRNA network. Node size in the figure represents the number of genes enriched in the pathway, and color represents the significance enriched.

LYS309 with supporting effects. The above results suggested that DB11093 has some affinity with SPRTN, KLHL28 and EPM2AIP1L and may become an inhibitor of the three targets. To the best of our knowledge, this is the first study that discovers a unique lncRNA-mRNA signature for MI detection and its pathogenic mechanism.

Our lncRNA-mRNA displayed excellent performance in MI diagnosis and prognosis. SPRTN gene has recently been found to be functional in translating DNA synthesis and preventing mutations [17, 18]. *In vivo* and *in vitro* characterization of identified mutations revealed that SPRTN plays an important role in the prevention of DNA replication stress during general DNA replication and in the regulation of replication-associated G2/M-checkpoints [19]. Mutations of SPRTN could lead to the early onset of hepatocellular carcinoma and genomic instability [19]. Moreover, the absence of EPM2AIP1 in mice impairs the allosteric activation of GS by glucose 6-phosphate, reduces liver glycogen synthesis, increases liver fat, and promotes liver insulin resistance [20]. LncRNA RP11-847H18.2 and mRNA KLHL28 has not been previously reported. Compared with using

lncRNA and mRNA alone in diagnosing and predicting MI prognosis, the lncRNA-mRNA pair developed in this study showed a higher discriminatory power, demonstrating a highly effectiveness of serving as a diagnostic and prognostic tool during MI detection.

Our functional enrichment analysis study revealed that the lncRNA-mRNA network is associated with autophagy and insulin resistance. Previous reports detected a sharp increase in autophagy during the reperfusion phase of cardiac ischemia [21, 22]. Kanamori et al. also confirmed that autophagy activity increases in the subacute and chronic phases of heart ischemia in MI mouse models [23, 24]. Noticeably, autophagy activity plays an even more critical role in the margin of infarction than in remote areas of the myocardium [25]. Insulin resistance is considered to be an adverse metabolic and a key risk factor for developing cardiovascular diseases [13], and insulin receptor substrates (IRS) are key modulators of insulin signal transduction pathways in the heart. After the activation of insulin receptor (INSR), IRS will promote the translocation of glucose transporter 4 (SLC2A4, also known as GLUT4) and initiates the cell survival

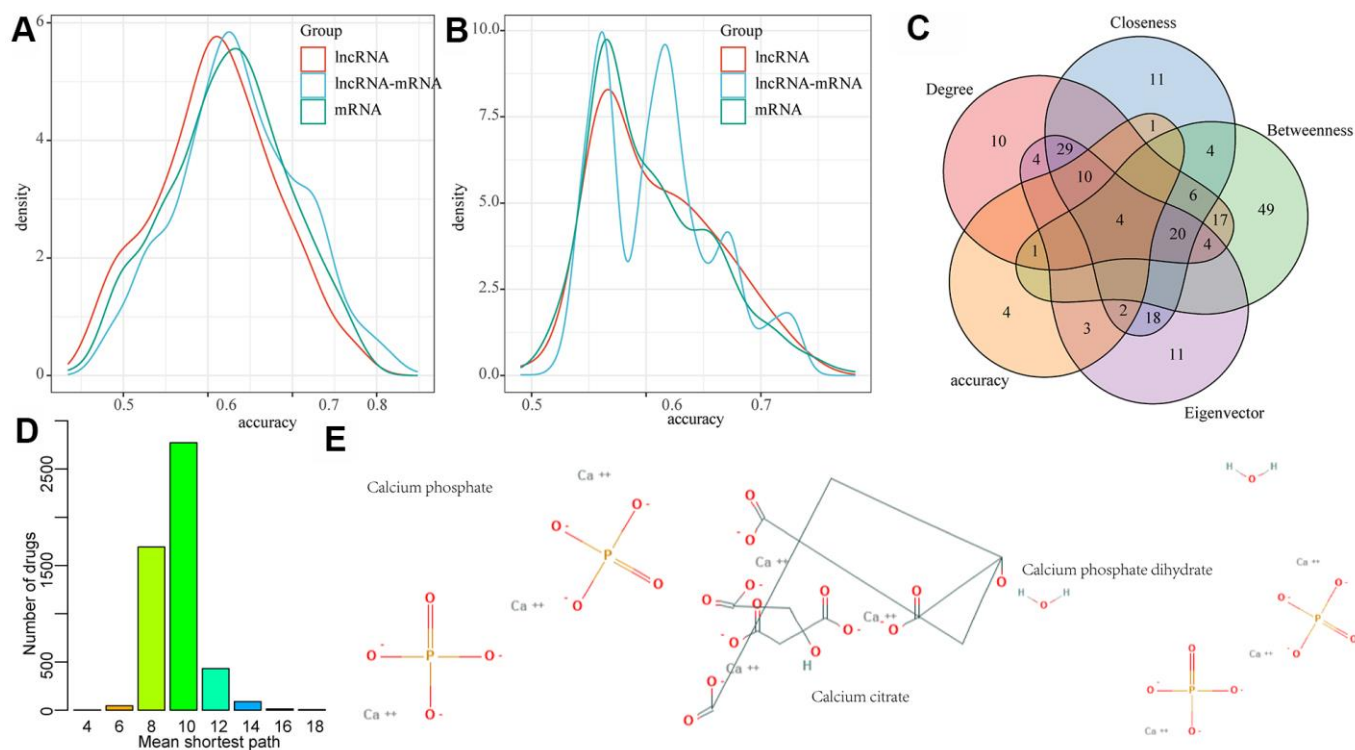


Figure 6. LncRNA-mRNA network mining was used to identify MI diagnostic and prognostic biomarkers. (A) The distribution of diagnostic accuracy of each lncRNA, mRNA, lncRNA-mRNA pair in the network for MI; (B) The distribution of prognostic of each lncRNA, mRNA, lncRNA-mRNA pair in the network for MI; (C) The intersection of candidate RNA molecules and nodes with network degree, medium centrality, near centrality and eigenvector centrality of top 10% in the network; (D) Mean shortest path distribution of drugs to MI gene markers. (E) 2D structures of three potential MI drug molecules.

pathway of cardiomyocytes [14]. The reduced insulin sensitivity in ischemic myocardium contributes to defective IRS1 function [15, 16]. These results indicate that the lncRNA and mRNAs in the network developed by this study may be involved in the development of MI through regulating autophagy and insulin resistance.

There are still some limitations in this study. For example, our sample size was relatively small for the evaluation of the correlation between the lncRNA-mRNA pair and the severity of MI or long-term clinical results. Thus, further basic research is required to verify the accuracy and clinical applicability of the lncRNA-mRNA pair in MI detection.

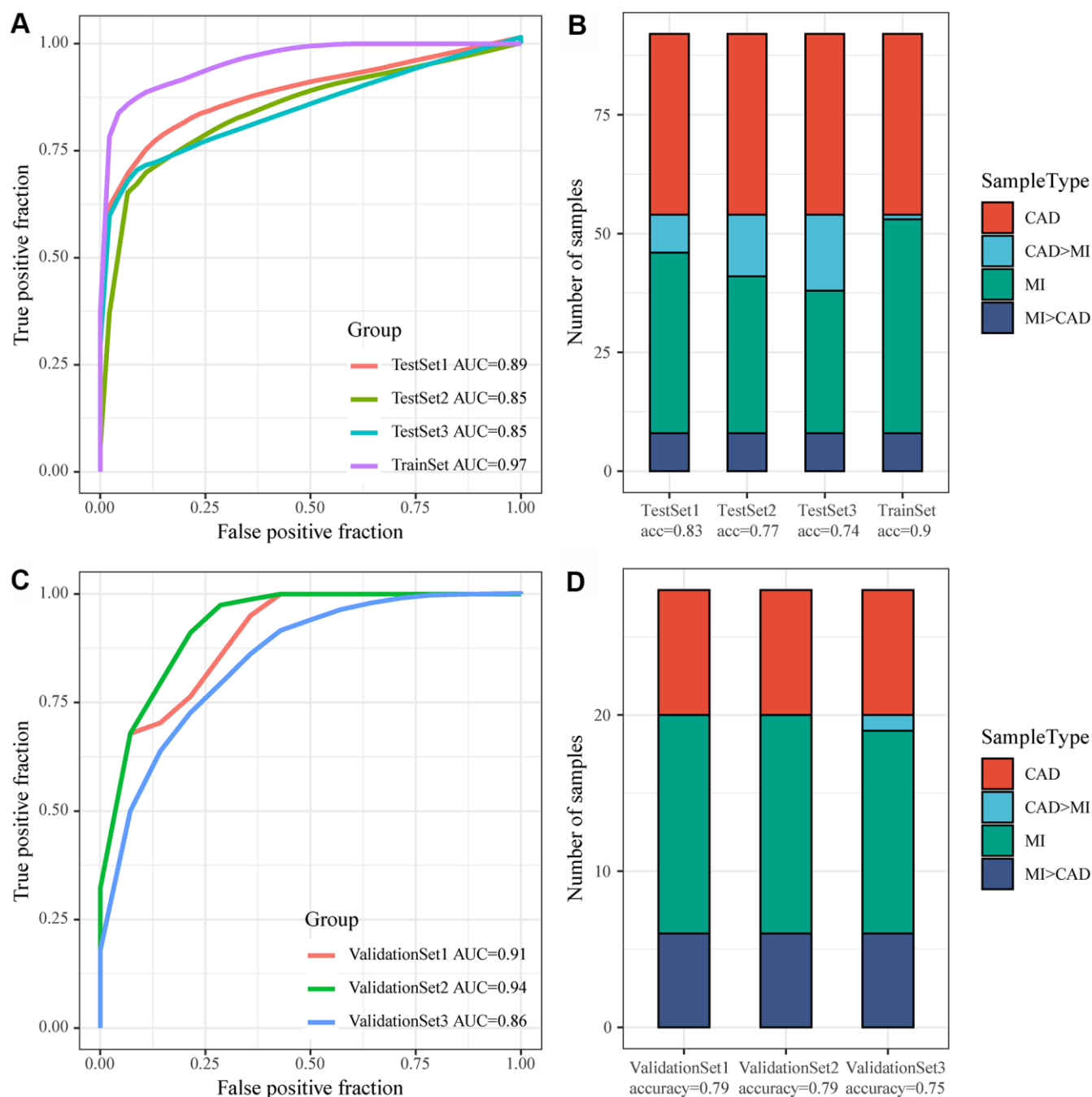


Figure 7. Construction of the MI diagnostic model and testing of the model. (A) ROC curve of the diagnostic model for classification of samples in the training dataset and internal validation dataset; **(B)** Classification accuracy of the diagnostic model on samples in the training set and in the internal validation set; **(C)** ROC curve of the diagnostic model for classification of samples in the validation dataset; **(D)** Classification accuracy of the diagnostic model on samples in the training set and in the validation dataset.

CONCLUSIONS

In conclusion, this study was the first to identify a unique gene signature (lncRNA RP11-847H18.2, mRNA KLHL28, SPRTN and EPM2AIP1), which allows an early detection of human MI. The gene signature specifically targets autophagy and insulin resistance, and may be involved in pathologies of MI.

MATERIALS AND METHODS

The experiment was conducted briefly as follows: data collection, co-expression module identification, time series analysis, enrichment analysis, feature selection, followed by classifier construction and validation.

RNA expression spectrum

GSE59867 expression profile data set of MI patients was acquired from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) [26] on the platform of Affymetrix Human Gene 1.0 ST Array. The dataset consisted of 456 samples, including 111 PATIENTS with ST-segment elevation MI and 46 patients with stable CAD without previous history of MI. The corresponding mRNA expressions were obtained from peripheral blood mononuclear cells at four time points, specifically on the day of admission (day 1, MI MI_S1), at discharge (4 to 6 days after MI, MI_S2), 1 months after MI (MI_S3) and 6 months after MI (MI_S4). The expression profiles of 73 out of 111 patients were detected at the four time points, so their expression profiles and 46 CAD samples served as controls in this study. In addition, we also downloaded the expression spectrum from the same platform dataset GSE62646 [27]. In this dataset, a total of 98 samples contained 28 patients with ST-segment elevation MI and 14 patients with stable CAD without MI history. Corresponding expression profiles were obtained from peripheral blood mononuclear cells at the day of admission (day 1 of MI, MI_S1), at discharge (4-6 days after MI, MI_S2) and 6 months after MI (MI_S4).

Probe sequences of GSE59867 and GSE62646 datasets were first aligned to the genome (GRCH38.p13) by chip reannotation to obtain probe mapping transcript IDs, and each transcript cluster was assigned to the Ensembl gene ID. Then, for transcription clusters with Ensembl gene ID, "LincRNA", "sense_intronic", "sense_overlapping", "antisense", "processed_transcript", "3_prime_overlapping_ncrna", "antisense_RNA", "TEC", and "bidirectional_promoter_lncRNA" were retained, but only "non_coding" cluster was considered as a lncRNA [28]. Finally, 1138 lncRNAs were collected after the removal of redundant transcripts. In addition, the cluster with annotation type "protein_coding" was

retained, and was considered as coding genes, here, we obtained a total of 14,099 coding genes.

The original data of GSE59867 and GSE62646 were processed by R Software package AFFy [29], and the expression matrix of probe was obtained by using RMA standardization. The probe was mapped to the gene, and the median value was taken as the expression value of the gene when multiple probes were mapped to the same gene. In this way, the expression matrixes of genes and lncRNAs were finally set up.

Expression analysis of lncRNAs and mRNAs, and the construction of weighted co-expression network

To better identify disease-related genes and lncRNAs, the expression profiles of lncRNAs and genes were combined for establishing a weighted co-expression module. Specifically, RNA expression data profiles of genes and lncRNAs were first tested to verify whether the samples, genes, and lncRNAs were qualified. Then, weighted gene co-expression network analysis (WGCNA) [30] package in R was applied to construct a scale-free co-expression network for the genes and lncRNAs, and the Pearson's correlation matrices and average linkage method were both performed for pair-wise testing. Then, a weighted adjacency matrix was constructed using a power function $A_{mn} = |C_{mn}|^\beta$ (C_{mn} = Pearson's correlation between gene/lncRNA m and gene/lncRNA n; A_{mn} = adjacency between gene/lncRNA m and gene/lncRNA n). β served as a soft-thresholding parameter that emphasizes strong correlations between gene, lncRNAs and removes weak correlations. Topological overlap matrix (TOM), which measures the connectivity of a gene/lncRNA to the network, is defined by the sum of a gene adjacency with all other gene/lncRNAs for network gene/lncRNA ratio. After deciding the power of β , the adjacency was converted into a TOM, and the corresponding dissimilarity (1-TOM) was calculated. Gene and lncRNAs with similar expression profiles were accordingly classified into corresponding modules, and average-linkage hierarchical clustering was conducted according to the TOM-based dissimilarity measure with a minimum size (gene/lncRNA group) of 30 for the genes/lncRNAs dendrograms. To further analyze the module, we calculated the dissimilarity of module Eigen of genes/lncRNAs, decided a cut line for module dendrogram and some modules were merged.

Identification of disease-related co-expression modules and lncRNA/miRNA clusters with consistent expression trends during MI progression

Co-expression modules eigengenes (MEs) were considered as the main components in the principal

component analysis of each module, and the expression patterns of all the genes and lncRNAs can be generalized into an expression profile with a single characteristic RNA in a given module. Therefore, correlations between ME and MI features at different time points were computed to identify MI-related co-expression modules. The R package "Mfuzz [31] was used to detect lncRNA and miRNA clusters with consistent expression trends during MI progression.

Functional enrichment analyses

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using the R package clusterProfiler [32] for genes to identify over-represented GO terms in three categories (biological processes, molecular function and cellular component) and KEGG pathway. For both analyses, $p < 0.05$ was considered to denote a statistical significance.

Regulatory interactions between miRNA-mRNAs and miRNA-lncRNAs

A total of 416312 non-redundant miRNA-mRNA interactions were obtained by acquiring all the regulatory relationships of miRNA-mRNA from the miRanda [33], miRTarBase [34], TargetScan [35], and starBase [36] databases. By predicting the miRNA-lncRNA interactions on the starBase [36] and miRcode [37] databases, 295,601 non-redundant miRNA-lncRNA relationships were retained.

Identification of MI-associated lncRNA-mRNA regulatory networks

Based on the ceRNA hypothesis [38, 39], a candidate lncRNA or mRNA is determined if it satisfies all of the following conditions: (1) the miRNA shared by mRNA and lncRNA is significantly enriched ($FDR < 0.01$ as determined by hypergeometric test); (2) the mRNA-lncRNA is significantly enriched in the same disease-associated co-expression module or in the same expression pattern (Cytoscope for visualization) [40]. Next, the topological properties of the network and the distribution of lncRNA-mRNA correlations in the network were analyzed to select lncRNA-mRNA pairs with correlations greater than 0.3 as candidates for MI correlation examination.

Network construction of gene markers and drug targets

To investigate the potential drug effects of these RNA markers, we identified a total of 16,196 drug-gene interactions from the DrugBank V5.1.7 database [41].

These drug target genes and RNA marker genes were co-mapped to the String V11.0 database (<https://string-db.org/>) [3] to obtain gene interaction information for the construction of a drug-gene-MI marker network. The shortest path from each drug to MI marker gene in the network was calculated, and the average shortest path of drug to MI marker was determined as the treatment candidate.

Construction of MI diagnostic prediction model and evaluation of model prediction performance

As a supervised learning model of machine learning algorithms, support vector machine (SVM) analyzes data and identifies patterns. A support vector mechanism creates a hyperplane for classification and regression in high or infinite dimensional space. MI gene markers and lncRNA markers were used to construct a diagnostic prediction model based on SVM [42] classification for the prediction on the MI and CAD samples. Given a set of training samples and each tag belongs to two categories, a SVM training algorithm establishes a model and assigns new instances to one category or another to allow a non-probabilistic binary linear classification. GSE59867 was taken as Train dataset, in order to balance the proportion between the control group and the disease group, we randomly selected 46 samples from 73 patients in the MI_S1 and 46 samples from the control group to form a queue as training set TrainSet. 46 samples were extracted from MI_S2, MI_S3 and MI_S4 and formed three data sets (TestSet1, TestSet2 and TestSet3, respectively) as the internal validation sets for the control group. Similarly, GSE62646 was regarded as the external validation set, and the three validation datasets (ValidationSet1, ValidationSet2 and ValidationSet3, respectively) were composed of the samples with the same proportion as the control group. Specifically, there were 14 control samples from the MI samples corresponding to the three follow-up time points in the GSE62646 dataset. The model was constructed in the training dataset, and its classification performance was examined by using the ten-fold cross-validation method. The established model was then used to predict the samples in the validated dataset. The predictive capability, predictive sensitivity and specificity of the model to MI were analyzed and evaluated by the area under the ROC curve (AUC).

AUTHOR CONTRIBUTIONS

All authors participated in the design, interpretation of the studies and analysis of the data and review of the manuscript; XCZ, GFW conducted conception and design of the research. XCZ, ZQC, and RQN conducted experiments. XCZ, JBZ performed acquisition of data. ZQC and JS analyzed data. CY contributed to the

statistical analysis. X CZ drafted the manuscript. GFW revised manuscript for important intellectual content. All authors participated in the design, interpretation of the studies and analysis of the data and review of the manuscript.

CONFLICTS OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

FUNDING

This work is supported by National Natural Science Foundation of China [Grant No. 819770367, 81670417, 81500361 and 81370389], National Key R&D Program of the Ministry of Science and Technology of China (Grant No. 2020YFC2004400), Science and Technology Planning Project of Shenzhen Municipality (Grant No. YJ20180306174831458), Public Health Research Fund of Futian District, Shenzhen [FTWS2018071], Shenzhen Science and technology R&D Fund (Grant No. JCYJ20160608142215491) and Shenzhen Key Medical Discipline Construction Fund (Grant No. SZXK002).

REFERENCES

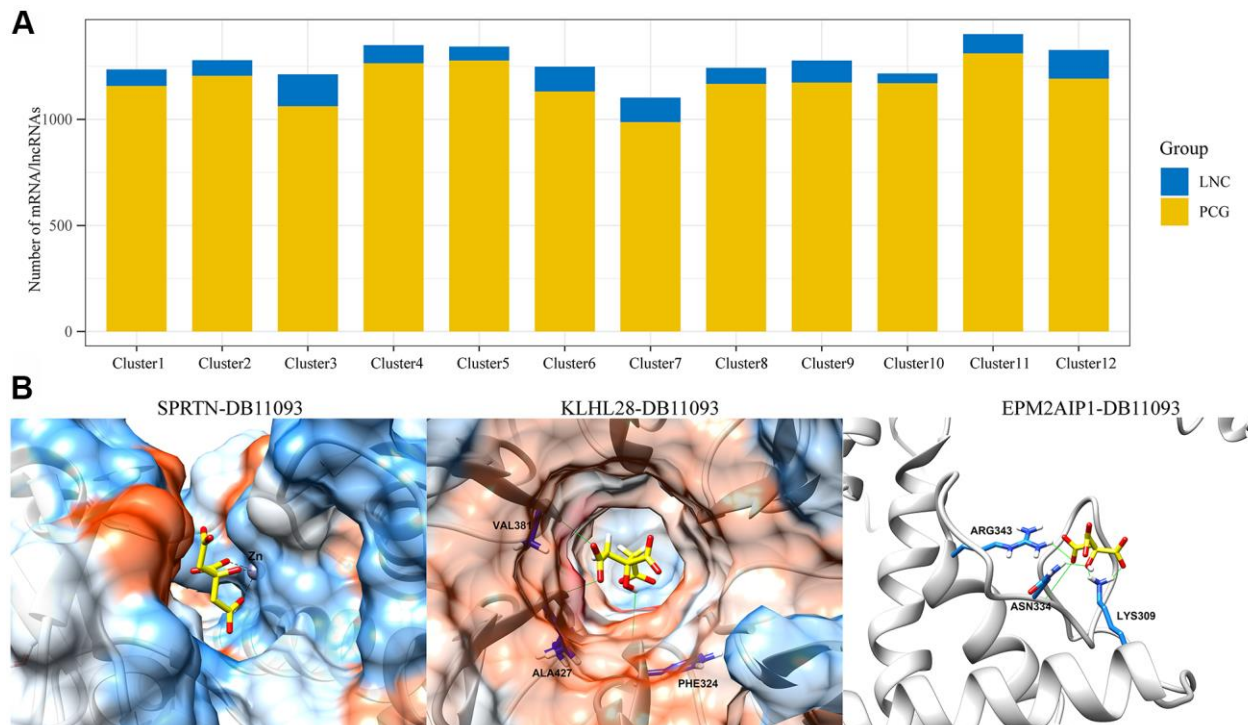
1. Lu L, Liu M, Sun R, Zheng Y, Zhang P. Myocardial infarction: symptoms and treatments. *Cell Biochem Biophys*. 2015; 72:865–67. <https://doi.org/10.1007/s12013-015-0553-4> PMID:25638347
2. Shinde AV, Frangogiannis NG. Fibroblasts in myocardial infarction: a role in inflammation and repair. *J Mol Cell Cardiol*. 2014; 70:74–82. <https://doi.org/10.1016/j.yjmcc.2013.11.015> PMID:24321195
3. Bhan A, Soleimani M, Mandal SS. Long noncoding RNA and cancer: a new paradigm. *Cancer Res*. 2017; 77:3965–81. <https://doi.org/10.1158/0008-5472.CAN-16-2634> PMID:28701486
4. Lozano-Vidal N, Bink DI, Boon RA. Long noncoding RNA in cardiac aging and disease. *J Mol Cell Biol*. 2019; 11:860–67. <https://doi.org/10.1093/jmcb/mjz046> PMID:31152659
5. Vausort M, Wagner DR, Devaux Y. Long noncoding RNAs in patients with acute myocardial infarction. *Circ Res*. 2014; 115:668–77. <https://doi.org/10.1161/CIRCRESAHA.115.303836> PMID:25035150
6. Li X, Dai Y, Yan S, Shi Y, Han B, Li J, Cha L, Mu J. Down-regulation of lncRNA KCNQ1OT1 protects against myocardial ischemia/reperfusion injury following acute myocardial infarction. *Biochem Biophys Res Commun*. 2017; 491:1026–33. <https://doi.org/10.1016/j.bbrc.2017.08.005> PMID:28780351
7. Li Y, Wang K, Wei Y, Yao Q, Zhang Q, Qu H, Zhu G. lncRNA-MIAT regulates cell biological behaviors in gastric cancer through a mechanism involving the miR-29a-3p/HDAC4 axis. *Oncol Rep*. 2017; 38:3465–72. <https://doi.org/10.3892/or.2017.6020> PMID:29039602
8. Zhang Y, Jiao L, Sun L, Li Y, Gao Y, Xu C, Shao Y, Li M, Li C, Lu Y, Pan Z, Xuan L, Zhang Y, et al. lncRNA ZFAS1 as a SERCA2a inhibitor to cause intracellular Ca²⁺ overload and contractile dysfunction in a mouse model of myocardial infarction. *Circ Res*. 2018; 122:1354–68. <https://doi.org/10.1161/CIRCRESAHA.117.312117> PMID:29475982
9. Zhang Y, Sun L, Xuan L, Pan Z, Li K, Liu S, Huang Y, Zhao X, Huang L, Wang Z, Hou Y, Li J, Tian Y, et al. Reciprocal changes of circulating long non-coding RNAs ZFAS1 and CDR1AS predict acute myocardial infarction. *Sci Rep*. 2016; 6:22384. <https://doi.org/10.1038/srep22384> PMID:26928231
10. Liu CY, Zhang YH, Li RB, Zhou LY, An T, Zhang RC, Zhai M, Huang Y, Yan KW, Dong YH, Ponnusamy M, Shan C, Xu S, et al. lncRNA CAIF inhibits autophagy and attenuates myocardial infarction by blocking p53-mediated myocardial transcription. *Nat Commun*. 2018; 9:29. <https://doi.org/10.1038/s41467-017-02280-y> PMID:29295976
11. Moe GW, Marín-García J. Role of cell death in the progression of heart failure. *Heart Fail Rev*. 2016; 21:157–67. <https://doi.org/10.1007/s10741-016-9532-0> PMID:26872675
12. Liu H, Liu P, Shi X, Yin D, Zhao J. NR4A2 protects cardiomyocytes against myocardial infarction injury by promoting autophagy. *Cell Death Discov*. 2018; 4:27. <https://doi.org/10.1038/s41420-017-0011-8> PMID:29531824
13. Ginsberg HN. Insulin resistance and cardiovascular disease. *J Clin Invest*. 2000; 106:453–58. <https://doi.org/10.1172/JCI10762> PMID:10953019
14. Guo CA, Guo S. Insulin receptor substrate signaling controls cardiac energy metabolism and heart failure. *J Endocrinol*. 2017; 233:R131–43. <https://doi.org/10.1530/JOE-16-0679> PMID:28381504

15. Ciccarelli M, Chuprun JK, Rengo G, Gao E, Wei Z, Peroutka RJ, Gold JI, Gumpert A, Chen M, Otis NJ, Dorn GW 2nd, Trimarco B, Iaccarino G, Koch WJ. G protein-coupled receptor kinase 2 activity impairs cardiac glucose uptake and promotes insulin resistance after myocardial ischemia. *Circulation*. 2011; 123:1953–62. <https://doi.org/10.1161/CIRCULATIONAHA.110.988642> PMID:21518983
16. Zhao YY, Xiao M, Zhang CL, Xie KQ, Zeng T. Associations between the tumor necrosis factor- α gene and interleukin-10 gene polymorphisms and risk of alcoholic liver disease: a meta-analysis. *Clin Res Hepatol Gastroenterol*. 2016; 40:428–39. <https://doi.org/10.1016/j.clinre.2015.10.007> PMID:26656007
17. Kim MS, Machida Y, Vashisht AA, Wohlschlegel JA, Pang YP, Machida YJ. Regulation of error-prone translesion synthesis by Spartan/C1orf124. *Nucleic Acids Res*. 2013; 41:1661–68. <https://doi.org/10.1093/nar/gks1267> PMID:23254330
18. Ghosal G, Leung JW, Nair BC, Fong KW, Chen J. Proliferating cell nuclear antigen (PCNA)-binding protein C1orf124 is a regulator of translesion synthesis. *J Biol Chem*. 2012; 287:34225–33. <https://doi.org/10.1074/jbc.M112.400135> PMID:22902628
19. Lessel D, Vaz B, Halder S, Lockhart PJ, Marinovic-Terzic I, Lopez-Mosqueda J, Philipp M, Sim JC, Smith KR, Oehler J, Cabrera E, Freire R, Pope K, et al. Mutations in SPRTN cause early onset hepatocellular carcinoma, genomic instability and progeroid features. *Nat Genet*. 2014; 46:1239–44. <https://doi.org/10.1038/ng.3103> PMID:25261934
20. Turnbull J, Tiberia E, Pereira S, Zhao X, Pencea N, Wheeler AL, Yu WQ, Ivovic A, Naranian T, Israelian N, Draginov A, Piliguian M, Frankland PW, et al. Deficiency of a glycogen synthase-associated protein, Epm2a1p, causes decreased glycogen synthesis and hepatic insulin resistance. *J Biol Chem*. 2013; 288:34627–37. <https://doi.org/10.1074/jbc.M113.483198> PMID:24142699
21. Huang Z, Han Z, Ye B, Dai Z, Shan P, Lu Z, Dai K, Wang C, Huang W. Berberine alleviates cardiac ischemia/reperfusion injury by inhibiting excessive autophagy in cardiomyocytes. *Eur J Pharmacol*. 2015; 762:1–10. <https://doi.org/10.1016/j.ejphar.2015.05.028> PMID:26004523
22. Ma X, Liu H, Foyil SR, Godar RJ, Weinheimer CJ, Diwan A. Autophagy is impaired in cardiac ischemia-reperfusion injury. *Autophagy*. 2012; 8:1394–96. <https://doi.org/10.4161/auto.21036> PMID:22889942
23. Kanamori H, Takemura G, Goto K, Maruyama R, Ono K, Nagao K, Tsujimoto A, Ogino A, Takeyama T, Kawaguchi T, Watanabe T, Kawasaki M, Fujiwara T, et al. Autophagy limits acute myocardial infarction induced by permanent coronary artery occlusion. *Am J Physiol Heart Circ Physiol*. 2011; 300:H2261–71. <https://doi.org/10.1152/ajpheart.01056.2010> PMID:21421825
24. Kanamori H, Takemura G, Goto K, Maruyama R, Tsujimoto A, Ogino A, Takeyama T, Kawaguchi T, Watanabe T, Fujiwara T, Fujiwara H, Seishima M, Minatoguchi S. The role of autophagy emerging in postinfarction cardiac remodeling. *Cardiovasc Res*. 2011; 91:330–39. <https://doi.org/10.1093/cvr/cvr073> PMID:21406597
25. Wang X, Guo Z, Ding Z, Mehta JL. Inflammation, autophagy, and apoptosis after myocardial infarction. *J Am Heart Assoc*. 2018; 7:e008024. <https://doi.org/10.1161/JAHA.117.008024> PMID:29680826
26. Maciejak A, Kiliszek M, Michalak M, Tulacz D, Opolski G, Matlak K, Dobrzycki S, Segiet A, Gora M, Burzynska B. Gene expression profiling reveals potential prognostic biomarkers associated with the progression of heart failure. *Genome Med*. 2015; 7:26. <https://doi.org/10.1186/s13073-015-0149-z> PMID:25984239
27. Kiliszek M, Burzynska B, Michalak M, Gora M, Winkler A, Maciejak A, Leszczynska A, Gajda E, Kochanowski J, Opolski G. Altered gene expression pattern in peripheral blood mononuclear cells in patients with acute myocardial infarction. *PLoS One*. 2012; 7:e50054. <https://doi.org/10.1371/journal.pone.0050054> PMID:23185530
28. Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, et al, and ENCODE Project Consortium, and Children’s Hospital Oakland Research Institute. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816. <https://doi.org/10.1038/nature05874> PMID:17571346
29. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy—analysis of affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004; 20:307–15. <https://doi.org/10.1093/bioinformatics/btg405> PMID:14960456

30. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9:559.
<https://doi.org/10.1186/1471-2105-9-559>
PMID:[19114008](https://pubmed.ncbi.nlm.nih.gov/19114008/)
31. Kumar L, E Futschik M. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics*. 2007; 2:5–7.
<https://doi.org/10.6026/97320630002005>
PMID:[18084642](https://pubmed.ncbi.nlm.nih.gov/18084642/)
32. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284–87.
<https://doi.org/10.1089/omi.2011.0118>
PMID:[22455463](https://pubmed.ncbi.nlm.nih.gov/22455463/)
33. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*. 2006; 126:1203–17.
<https://doi.org/10.1016/j.cell.2006.07.031>
PMID:[16990141](https://pubmed.ncbi.nlm.nih.gov/16990141/)
34. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, Tang Y, Chen YG, Jin CN, Yu Y, Xu JT, Li YM, Cai XX, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res*. 2020; 48:D148–54.
<https://doi.org/10.1093/nar/gkz896> PMID:[31647101](https://pubmed.ncbi.nlm.nih.gov/31647101/)
35. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005; 120:15–20.
<https://doi.org/10.1016/j.cell.2004.12.035>
PMID:[15652477](https://pubmed.ncbi.nlm.nih.gov/15652477/)
36. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-seq data. *Nucleic Acids Res*. 2014; 42:D92–97.
<https://doi.org/10.1093/nar/gkt1248>
PMID:[24297251](https://pubmed.ncbi.nlm.nih.gov/24297251/)
37. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*. 2012; 28:2062–63.
<https://doi.org/10.1093/bioinformatics/bts344>
PMID:[22718787](https://pubmed.ncbi.nlm.nih.gov/22718787/)
38. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011; 146:353–58.
<https://doi.org/10.1016/j.cell.2011.07.014>
PMID:[21802130](https://pubmed.ncbi.nlm.nih.gov/21802130/)
39. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature*. 2014; 505:344–52.
<https://doi.org/10.1038/nature12986>
PMID:[24429633](https://pubmed.ncbi.nlm.nih.gov/24429633/)
40. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–504.
<https://doi.org/10.1101/gr.1239303>
PMID:[14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
41. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018; 46:D1074–D1082.
<https://doi.org/10.1093/nar/gkx1037>
PMID:[29126136](https://pubmed.ncbi.nlm.nih.gov/29126136/)
42. Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinformatics*. 2018; 19:432.
<https://doi.org/10.1186/s12859-018-2451-4>
PMID:[30453885](https://pubmed.ncbi.nlm.nih.gov/30453885/)

SUPPLEMENTARY MATERIALS

Supplementary Figure



Supplementary Figure 1. (A) Distribution statistics of genes and lncRNAs under various co-expression patterns. (B) Diagram of interaction between KLHL28, SPRNTN and EPM2AIP1 and the drug DB11093, Coordination bonds are shown as the black dotted line and hydrogen bonds as the green solid line. DB11093 is shown as a yellow stick and the color of the heteroatoms on it is shown as an element. In addition, hydrophobic surfaces were added in images (A, B) to better observe the binding sites of the ligand proteins.